

**Code States AIB(AI Bootcamp)   
수강생 프로젝트 기획서**

|  |  |
| --- | --- |
| 수강생 이름 : | 오상훈 |
| 시작일 : | 2021. 10. 08 |
| 프로젝트 : | SNS 리뷰데이터 분석 |

|  |
| --- |
| **1. 프로젝트 주제** |
| SNS 관광리뷰 자연어 활용 분석 |
| **2. 프로젝트 목적** |
| 경기도 관광지 SNS 리뷰데이터 분석을 통해 방문객 평가 분석을 통해 개선점 및 관광활용 방안 도출 |
| **3. 프로젝트 가설, 예상결과** |
| 리뷰데이터를 토대로 인기/비인기 관광지역과 빈도수 높은 단어 감성분석 |
| **4. 프로젝트 분석방법** |
| * 데이터셋 전처리(Pandas, 정규식 활용) * 토큰화(전체 리뷰, 관광지별 리뷰, 불용어 처리) * 데이터 시각화(단어 빈도수 순위, 워드클라우드) * 분석된 자료를 토대로 관광지 개선점 및 관광활용 방안 도출 |
| **5. 프로젝트 언어, 환경** |
| * 주 프로그래밍 언어: Python(파이썬) * 작성 프로그램: Jupyter Notebook(주피터 노트북) * 주 활용 라이브러리: Pandas, KoNlPy(코엔엘파이)-Okt, WordCloud |
| **6. 프로젝트 결과물** |
| * 프로젝트 A-Z 구글 프레젠테이션 발표 영상 * 프로젝트 소개 및 관련 파일Git Hub 페이지 * 분석 및 모델링을 담은 Jupyter Notebook 파일 & PPT 파일 |

|  |  |
| --- | --- |
| **7. 프로젝트 데이터** | |
| **데이터 이름**  (파일명) | **데이터 설명**  (어떠한 데이터 인지 최대한 설명해 주시기 바랍니다) |
| 데이터 1  (01\_Raw\_data) | 경기지역 관광지 SNS리뷰데이터 |
| 데이터 2  (02\_Preprocessing) | 1차 전처리 된 데이터(한글 데이터만 추출, 특수문자 제거 등) |
| 데이터 3  (03~05\_StopWords) | 불용어 사전(자체 제작, 1차~3차) |
| 데이터 4  (06\_all\_token\_data) | 토큰화 된 경기지역 관광지 SNS리뷰데이터 |

|  |
| --- |
| **8. 프로젝트 구현 예시** |
|  |